

## Amendments to the Claims

Claim 1 (currently amended): A system for disambiguating speech input using one of voice mode interaction, visual mode interaction, or a combination of voice mode interaction and visual mode interaction with an application comprising:

a speech disambiguation mechanism resident on one of an end user device and a remote server, and accessed through said end user device possessing multimodal user interfaces, said speech disambiguation mechanism comprising: [[:]]

an options and parameters component for receiving and storing user parameters and receiving application parameters for controlling the speech disambiguation mechanism, wherein the speech disambiguation mechanism is controlled by parameters set by the user and parameters set by the application, and wherein the parameters include confidence thresholds governing unambiguous recognition and close matches;

a speech recognition component that receives recorded audio, speech input or a combination of the recorded audio and the speech input through one of said multimodal user interfaces, and generates:

a plurality of tokens corresponding to disambiguated words for presentation to the user; and

for each of the ~~one or more~~ tokens, a confidence value indicative of the likelihood that a given token correctly represents the speech input;

a selection component that identifies, according to a selection algorithm, two or more of the tokens to be presented to the user;

one or more disambiguation components directing one or more of said multimodal user interfaces to ~~[[that]]~~ present the alternatives to the user in one of voice mode, visual mode, or a combination of the voice mode and the visual mode, and directing the multimodal user interfaces to ~~receives~~ receive an alternative selected by the user in one of the voice mode, the visual mode, or a combination of the voice mode and the visual mode; and

an output interface ~~that presents~~ for communicating the selected alternative without translation of the speech input to the application as input.

Claims 2 -3 (canceled).

Claim 4 (original): The system of claim 1, wherein the one or more disambiguation components perform said interaction by presenting the user with alternatives in a visual mode, and by receiving the user's selection in a visual mode.

Claim 5 (original): The system of claim 4, wherein the disambiguation components present the alternatives to the user in a visual form and allow the user to select from among the alternatives using a voice input.

Claim 6 (canceled).

Claim 7 (original): The system of claim 1, wherein the selection component filters the one or more tokens according to a set of parameters.

Claim 8 (original): The system of claim 7, wherein the set of parameters is user specified.

Claims 9 - 10 (canceled).

Claim 11 (currently amended): A method of processing speech input using one of voice mode interaction, visual mode interaction, or a combination of voice mode and visual mode interaction with an application comprising:

a speech disambiguation mechanism, wherein said speech disambiguation mechanism is resident on one of an end user device and a remote server, and accessed through said end user device possessing multimodal user interfaces;

receiving and storing user parameters and receiving application parameters for controlling the speech disambiguation mechanism, wherein both the user and the application can set the parameters to control said speech disambiguation mechanism, and wherein the parameters include confidence thresholds governing unambiguous recognition and close matches;

receiving a speech input from the user through one of said multimodal user interfaces;

determining whether the speech input is ambiguous;

if the speech input is not ambiguous, communicating a token representative of the speech input to ~~[[an]]~~ the application as input to the application; and

if the speech input is ambiguous;

selecting two or more tokens and presenting the tokens as alternatives to the user;

directing the multimodal user interfaces to present ~~presenting~~ the alternatives to the user in one of voice mode, visual mode, or a combination of the voice mode and the visual mode, and receiving to present a selection of an alternative from the user from the plurality of

alternatives presented to the user in one of the voice mode, the visual mode, or a combination of the voice mode and the visual mode; and

communicating the selected alternative without translation of the speech input as input to the application.

Claim 12 (original): The method of claim 11, where the interaction comprises the concurrent use of said visual mode and said voice mode.

Claim 13 (original): The method of claim 12, wherein the interaction comprises the user selecting from among the plural alternatives using a combination of speech and visual-based input.

Claim 14 (original): The method of claim 11, wherein the interaction comprises the user selecting from among the plural alternatives using visual input.

Claim 15 (new): The system of claim 1 further comprises a communication network, wherein the options and parameters component, the speech recognition component, the selection component, the one or more disambiguation components, and the output interface of the speech disambiguation mechanism are distributed on said communication network.

Claim 16 (new): A method of processing speech input using one of voice mode interaction, visual mode interaction, or a combination of voice mode and visual mode interaction with an application comprising:

a speech disambiguation mechanism, wherein said speech disambiguation mechanism is resident on a remote server, and accessed over a communication network using an end user device possessing multimodal user interfaces;

receiving and storing user parameters and receiving application parameters for controlling the speech disambiguation mechanism, wherein both the user and the application set the parameters to control said speech disambiguation mechanism, and wherein the parameters include confidence thresholds governing unambiguous recognition and close matches;

receiving a speech input from the user through one of said multimodal user interfaces;

determining whether the speech input is ambiguous;

if the speech input is not ambiguous, communicating a token representative of the speech input to the application as input to the application; and

if the speech input is ambiguous;

selecting two or more tokens and presenting the tokens as alternatives to the user;

directing the multimodal user interfaces to present the alternatives to the user in one of voice mode, visual mode, or a combination of the voice mode and the visual mode, and to present a selection of an alternative from the user from the plurality of alternatives presented to the user in one of the voice mode, the visual mode, or a combination of the voice mode and the visual mode; and

communicating the selected alternative without translation of the speech input as input to the application.

Claim 17 (new): The method of claim 16, where the interaction comprises the concurrent use of said visual mode and said voice mode.

Claim 18 (new): The method of claim 17, wherein the interaction comprises the user selecting from among the plural alternatives using a combination of speech and visual-based input.

Claim 19 (new): The method of claim 16, wherein the interaction comprises the user selecting from among the plural alternatives using visual input.